Dec 11th, 12:00 AM

Predicting Yield Before Harvest: How Does the USDA Forecast Corn and Soybean Yield?

Greg Thessen
U.S. Department of Agriculture

Predicting yield before harvest: How does the USDA forecast corn and soybean yield?

Greg Thessen, Director, Iowa Field Office, USDA-NASS

Introduction

Crop production forecasts have two components--acres to be harvested and expected yield per acre. For example, preliminary corn and soybean acreage estimates are made using data obtained from a survey of farmers conducted during the first 2 weeks in June. Expected corn and soybean yields are obtained monthly, August through November, from two different types of yield surveys. Data from the yield surveys reflect conditions as of the first of the month, as data are collected during the last week of the previous month and the first 2 or 3 days of the current month.

Crop production forecasts are based on conditions as of the survey reference date and projected assuming normal conditions for the remainder of the season. For example, the assumption of "normal conditions" is that temperatures and precipitation will be at historic averages for the remainder of the season. It is assumed that the first killing frost will occur on the historic average date. The crop maturity and conditions at the reference date are evaluated against the time remaining until the expected frost--if one third of the crop will not reach maturity until the frost date has passed; it is assumed that some frost damage will result. Long-range weather projections are not used as an indicator for final yield.

The reference point for crop forecast surveys is the first of the month, which is also usually close to the mid-point of data collection. Both grower-reported average yields and objective-measurement modeled yields contain a measurable forecast error based on the historic difference between these survey estimates and the final end-of-season yield. The review process followed to develop the monthly yield forecasts involves evaluating the relative ranges of the forecast errors of the grower yields and the objective measurement yields and the degree to which they overlap.

When NASS states as policy that it is forecasting based on conditions as of the first of the month, it is saying that it will establish yields within the range of the survey estimates.

When forecasting crop yields, NASS does not attempt to predict future weather conditions. Long-range weather forecasts are not used in any forecast models. To the extent that conditions depart from normal, the forecasts also will fluctuate. Procedures used to prepare acreage estimates and yield forecasts are discussed in the following sections.

Base for acreage planted and to be harvested

The largest single survey NASS conducts each year is the June Agricultural Survey. During the first 2 weeks in June, about 2,400 interviewers contact over 125,000 farmers, either by telephone or in person, to obtain information on crop acreages, grain stocks, and livestock inventories. These producers are asked to report the acreage, by crop, that has either been planted or that they intend to plant, and the acreage they expect to harvest as grain. Data from this survey are used to estimate, among other things, total acres planted to corn, soybeans, and other crops regardless of the intended uses. Preliminary projections of acres to be harvested for grain or

soybeans, including seed, are also made using these data.

The sample design for this survey utilizes two different sampling frames. The area frame, which is essentially the entire land mass of the United States, ensures complete coverage of the U.S. farm population. The list frame, a list of known farmers and ranchers, does not provide complete coverage of all farms, but allows the use of more efficient data collection methods.

Sampling from the area frame is a multi-step process. First, all land in each State is classified into land use categories by intensity of cultivation using a variety of map products, satellite imagery, and computer software packages. These land use classifications range from intensively cultivated areas to marginally cultivated grazing areas to urban areas. The land in each use category is then divided into segments ranging from about 1 square mile in cultivated areas to 0.1 square mile in urban areas. This allows intensively cultivated land segments to be selected with a greater frequency than those in less intensively cultivated areas. Segments representing cultivated areas are selected at a rate of about 1 out of 125. Sample segments in land use classifications with decreasing amounts of cultivated land are selected at rates ranging from 1 out of 250 to 1 out of 500.

About 10,000 area segments are selected nationwide for the survey conducted each June with 452 of these segments located in Iowa. Using maps and aerial photos showing the exact location and boundaries of each sample segment, interviewers locate and interview every operator with land inside the segment boundaries to identify crops planted in each field, and to obtain livestock inventory information, and quantities of grain in storage.

Before sampling from the list, each farm is classified by various characteristics such as number of acres by crop. Large farms are sampled at high rates. For example, Iowa farms on the list with over 5,000 acres of cropland, or grain storage capacity exceeding 1 million bushels, are selected with certainty. Smaller farms are selected at rates of 1 out of 10 to 50.

About 75,000 farms across the United States are selected from the list to be surveyed during the same time period in June with about 2500 of these farms located in Iowa. Farmers on the list sample are asked to provide total acres planted for each crop on all the land they operate, and quantities of grain stored on their operation. Most of the data from this sample are collected by telephone interviewers.

Data from the area and list samples are combined using multiple-frame statistical methodology developed jointly by NASS and Iowa State University, which ensures that all land areas in the United States can be accounted for once and only once.

Generally, estimates of planted acres for corn and soybeans from the June Agricultural Survey are not changed until October. However, occasionally the planting season runs late and many fields are not yet planted with the intended crops at the time the June survey is conducted. When this happens, adjustments to planted area estimates may be made at the time of the first yield forecast in August. If a significant portion of the crops are not planted by the time the June Agricultural Survey is competed (like in Iowa for 2008), NASS may re-interview the June survey respondents during late July to determine what was actually planted. The preliminary projections for harvested acres may also be adjusted using data from the August yield surveys or in extreme cases from a special July re-interview survey. Any necessary changes to planted and harvested acreage estimates will be published in the August *Crop Production* report.

In October, NASS will review several data sources for corn and soybeans, including the farmer reported surveys, satellite imagery, and acreage data reported by producers to the Farm Service Agency (FSA) and may update the area planted and expected acres for harvest in the October *Crop Production* report.

Yield forecasts

A subsample of farmers who respond to the list portion of the June Agricultural Survey is selected to provide monthly crop yield forecasts. This provides a way to screen farmers so that only those currently growing the commodities of interest are contacted during the monthly surveys. This monthly Agricultural Yield Survey asks the sampled farmers to report what they expect their crops to yield before harvest and actual yields are obtained at harvest. All yield data for an individual report are weighted by the farm's crop acres for harvest.

Objective Yield Surveys are conducted monthly in states that contribute most heavily to total U.S. production of corn and soybeans. These surveys provide information for making forecasts and estimates of crop yields based on counts, measurements, and weights obtained from small plots in a random sample of fields. Sample corn and soybean fields are selected from those identified in the area-frame sample portion of the June Agricultural Survey. In Iowa, 330 corn fields and 240 soybean fields are selected. Observations within each selected field are made in two randomly located plots. Plots include two adjacent rows of predetermined length.

Harvested yield can be thought of as biological or gross yield minus harvest loss. Counts, measurements, and other observations from each sample plot are input to statistical models based on historical data to predict final number of fruit and final weight per fruit. A forecast of gross yield is calculated by multiplying these two components together and dividing by land area. Table 1 shows the forecast variables used to predict the gross yield components for each crop.

Table 1. Objective yield forecast variables for number of fruit and fruit weight.

Crop	Component	Forecast variable[1]
corn	ears	stalks
		ears and ear shoots
		ears with kernels
	ear weight	historic average
		length over husk
		kernel row length
		ear diameter
soybean	plants	plants
	pods per plant	main stem nodes
		lateral branches
		blooms, dried flowers and pods
		pods with beans
	pod weight	historical average
		pods with beans

[1] Variables measured are determined by stage of maturity.

Plant characteristics used as prediction variables change as the crop maturity progresses. At an early stage, plant counts may be the only data available for forecasting the number of mature fruit. As the crop matures, actual fruit counts can be used, and weights and measurements of the immature fruit are used to predict final weight per fruit.

The same plots are revisited each month until the crop is mature. At that time, the plots are harvested and final counts and weights are obtained. After the entire field has been harvested by the farmer, one-fourth of the sample fields are revisited and two more plots are laid out. The grain left on the ground in these plots is picked up and weighed to provide a measure of harvest loss.

The estimate of number of soybean pods per acre from the Objective Yield Survey is usually very consistent from month to month and accurate when the bloom period has ended. Pod count forecasts usually stabilize with the September survey.

Average pod weights prior to crop maturity are based on historical averages. In normal years, much of the soybean crop has matured by the October survey, so current-year pod weights are used. Corn objective yield survey forecasts are based on estimates of number of ears and average ear weight. The ear count forecasts are accurate early in the season. When the crop is late developing, the August projection of ears is based on a model using plant population. Historical average ear weights are used until ears are present to measure. Kernel row length models are then used to project ear weight until crop maturity.

Potential accuracy of each month's forecast for these crops is dependent on the crop maturity at the time of the forecast and future weather. When maturity lags normal patterns, number of pods, ears, etc., is based on number of plants and fruiting positions rather than actual number of fruit. Thus, when maturity lags, the forecasts become more variable because the expected number of fruit can differ from the final. However, the primary source of forecast error occurs when final end of season fruit weights differ from the historic average because fruit weight cannot be fully determined until crop maturity. A comparison of the August and October forecast to the final NASS estimate for corn and soybeans is shown in Figures 1 and 2, respectively.

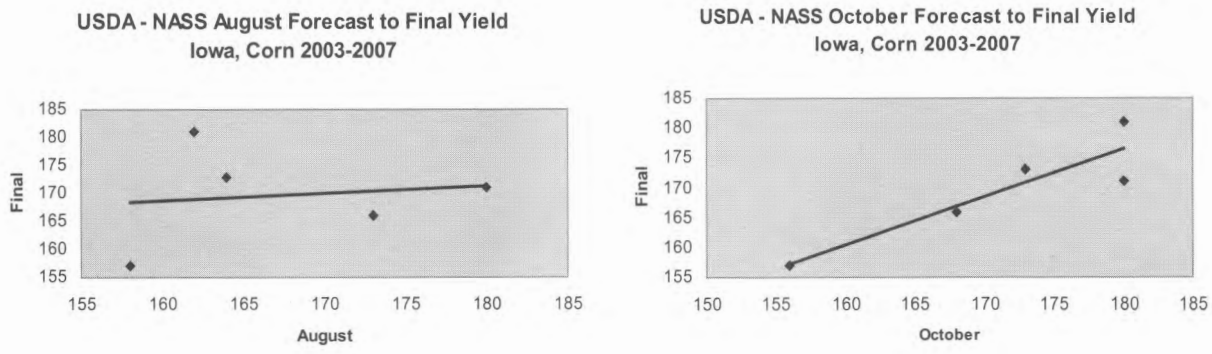

Figure 1

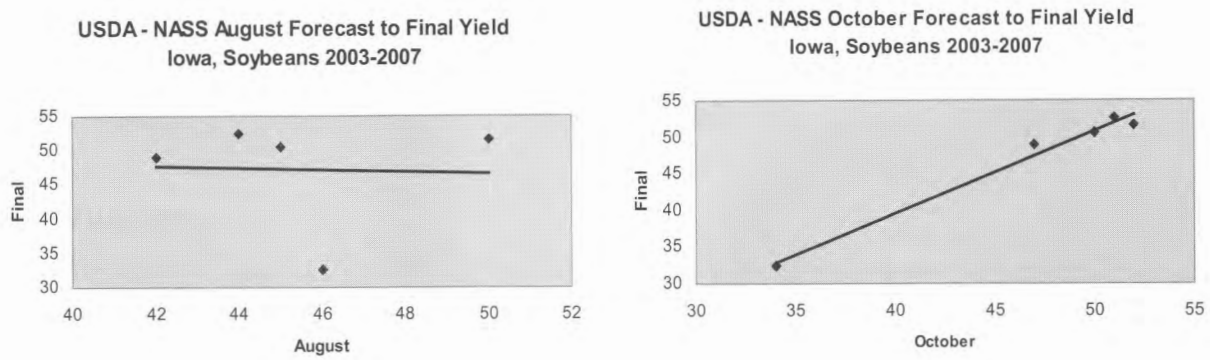

Figure 2

NASS will revise estimates of harvested acres if necessary during the forecast season when extreme weather events (flood, drought, hurricane) warrant and supporting data from a survey is available. Again, the goal is to make the production forecasts as accurate as possible. The production forecasts are based on projecting the acres that will be harvested and the final yield per harvested acre. If acres are lost during the forecast season because of weather or disease problems, those yields drop to zero, the acres are classified as planted but abandoned, and acres for harvest reduced. For this reason, it is possible for the production forecast to be reduced without a corresponding drop in forecast yield per acre. It is also possible for the yield per acre to increase during adverse periods if acres for harvest are abandoned and classified as not for harvest. Data on which to base changes in harvested acres come from the yield forecast surveys when sample fields are taken out of production or the operator reports acres no longer being considered for harvest.

USDA-NASS strives to provide the agricultural community with estimates and forecasts that are accurate, objective, reliable, and timely. This is accomplished by conducting surveys throughout the growing season and after harvest is complete. Security measures are used to prevent leaks of this market-sensitive information which is published in the *Crop Production* report each month and is available on the NASS website at **www.nass.usda.gov**.